IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: **Busche**	§	
	§	Group Art Unit: **3622**
Serial No. **09/879,491**	§	
	§	Examiner: **Lastra, Daniel**
Filed: **June 12, 2001**	§	
	§	
For: **Method and System for Predicting**	§	
Customer Behavior Based on Data	§	
Network Geography		

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

37945
PATENT TRADEMARK OFFICE
CUSTOMER NUMBER

APPEAL BRIEF (37 C.F.R. 41.37)

This brief is in furtherance of the Notice of Reinstatement of Appeal, filed in this case on August 31, 2007.

No fees are believed to be required. If, however, any fees are required, I authorize the Commissioner to charge these fees, which may be required to IBM Corporation Deposit Account No. 09-0457. No extension of time is believed to be necessary. If, however, an extension of time is required, the extension is requested, and I authorize the Commissioner to charge any fees for this extension to IBM Corporation Deposit Account No. 09-0457.

<u>REAL PARTY IN INTEREST</u>

The real party in interest in this appeal is the following party: International Business Machines Corporation of Armonk, New York.

RELATED APPEALS AND INTERFERENCES

With respect to other appeals or interferences that will directly affect, or be directly affected by, or have a bearing on the Board's decision in the pending appeal, there are no such appeals or interferences.

STATUS OF CLAIMS

A.   TOTAL NUMBER OF CLAIMS IN APPLICATION

Claims in the application are: 1-43

B.   STATUS OF ALL THE CLAIMS IN APPLICATION

1. Claims canceled: 9, 23 and 36
2. Claims withdrawn from consideration but not canceled: none
3. Claims pending: 1-8, 10-22, 24-35 and 37-43
4. Claims allowed: none
5. Claims rejected: 1-8, 10-22, 24-35 and 37-43
6. Claims objected to: none

C.   CLAIMS ON APPEAL

The claims on appeal are: 1-8, 10-22, 24-35 and 37-43

STATUS OF AMENDMENTS

No amendment after final rejection was filed for this case.

SUMMARY OF CLAIMED SUBJECT MATTER

Currently, when using artificial intelligence algorithms to discover patterns in behavior exhibited by customers, it is necessary to create training data sets where a predicted outcome is known as well as testing data sets where the predicted outcome is known to be able to validate the accuracy of a predictive algorithm. The predictive algorithm, for example, may be designed to predict a customer's propensity to respond to an offer or his propensity to buy a product. The data used to train and test the algorithm are selected using a random selection procedure, such as selecting data based upon a random number generator, or by some other means to insure that both the training data and test data sets are representative of the entire data population being evaluated. Tests of randomness of each of the attributes, e.g., the demographic information of the individuals, in the data sets can then be completed to see if they represent a randomly selected population.

While the above approach to selecting testing and training data sets may be suited for some applications, the purchasing behavior of customers is not only based on demographic and cyclographic information. Ease of access to various goods and services may also influence the customer's ultimate purchase patterns. That is, if a customer is able to obtain access to the goods and services more easily, the customer is typically more likely to engage in the purchase of such goods and services.

Today, customers are purchasing more and more goods and services over data networks, such as the Internet. In doing so, customers must often navigate a morass of web sites and web pages to ultimately arrive at the goods and services that they wish to purchase. This web sites and web pages that make up the data network are collectively referred to as the data network geography. Many times, a customer may become frustrated during this navigating of the data network geography and may abandon the endeavor. Other times, the customer may simply purchase goods and services from the first web site or web page that they locate that provides the goods and services without bothering to look at other web sites that may offer the same goods and services under different terms, such as pricing, incentives, and the like. Such influences on customer behavior by the data network geography are not taken into consideration when training

and using predictive algorithms to predict customer behavior. Thus, bias may be introduced into the test data, train data, or both the test and train data sets - making either one or both of them non-representative of the overall customer database.

Therefore, it would be beneficial to have a method and system for correlating a customer's effort in navigating a data network with the customer's purchase behavior; for predicting a customer's behavior based on the geography of the data network; and for evaluating the training of a predictive algorithm to determine if the training and testing data sets do not adequately take into consideration the influences of the data network geography on customer behavior.

A.   CLAIM 1 - INDEPENDENT

Claim 1 is directed to a data processing machine implemented method of selecting data sets for use with a predictive algorithm based on data network geographical information. A first statistical distribution of a training data set is generated. In addition, a second statistical distribution of a testing data set is generated. Both of these first and second statistical distributions are used to identify a discrepancy between the first statistical distribution and the second statistical distribution with respect to the data network geographical information by comparing the first statistical distribution and/or the second statistical distribution to a statistical distribution of a customer database in order to determine if the training data set and/or the testing data set are geographically representative of a customer population represented by the customer database. The selection of entries is modified in the training data set and/or the testing data set based on the discrepancy between the first statistical distribution and the second statistical distribution. This modified selection of entries is used by the predictive algorithm, thereby advantageously correlating a customer's effort in navigating a data network with the customer's purchase behavior; for predicting a customer's behavior based on the geography of the data network; and for evaluating the training of a predictive algorithm to determine if the training and testing data sets do not adequately take into consideration the influences of the data network geography on customer behavior (Specification page 15, line 21 – page 18, line 26; page 47, line 21 – page 48, line 22; Figure 6, all blocks).

B. CLAIM 15 – INDEPENDENT

Claim 15 is directed to an apparatus for selecting data sets for use with a predictive algorithm based on data network geographical information. The apparatus includes a statistical engine and a comparison engine coupled to the statistical engine. The statistical engine generates a first statistical distribution of a training data set and a second distribution of a testing data set. The comparison engine uses the first statistical distribution and the second distribution to identify a discrepancy between the first statistical distribution and the second distribution with respect to the data network geographical information by comparing the first statistical distribution and/or the second statistical distribution to a statistical distribution of a customer database to determine if the training data set and/or the testing data set are geographically representative of a customer population represented by the customer database. The comparison engine modifies the selection of entries in the training data set and/or the testing data set based on the discrepancy between the first statistical distribution and the second distribution. The comparison engine provides the modified selection of entries for use by the predictive algorithm. A predictive algorithm device uses this modified selection of entries and the predictive algorithm (Specification page 15, line 21 – page 18, line 26; page 47, line 21 – page 48, line 22; Figure 6, all blocks).

C. CLAIM 29 - INDEPENDENT

Claim 29 is directed to a computer program product in a computer readable medium. The computer program product includes instructions for enabling a data processing machine to select data sets for use with a predictive algorithm based on data network geographical information, including (1) instructions for generating a first statistical distribution of a training data set; (2) instructions for generating a second statistical distribution of a testing data set; (3) instructions for using the first statistical distribution and the second statistical distribution to identify a discrepancy between the first statistical distribution and the second statistical distribution with respect to the data network geographical information by comparing the first statistical distribution and/or the second statistical distribution to a statistical distribution of a customer database to determine if the training data set and/or the testing data set are geographically representative of a customer population represented by the customer database; (4) instructions for modifying selection of entries in the training data set and/or the testing data set based on the discrepancy between the first statistical distribution and the second statistical distribution; and (5)

instructions for using the modified selection of entries by the predictive algorithm (Specification page 15, line 21 – page 18, line 26; page 47, line 21 – page 48, line 22; Figure 6, all blocks).

D.   CLAIM 41 – INDEPENDENT

Claim 41 is directed to a data processing machine implemented method of predicting customer behavior based on data network geographical influences. Data network geographic information regarding a plurality of customers is obtained, where the data network geographic information includes frequency distributions of both (i) number of data network links between a customer geographical location and one or more web site data network geographical locations, and (ii) size of a click stream for arriving at the one or more web site data network geographical locations. A predictive algorithm is trained using the data network geographical information. This predictive algorithm is used to predict customer behavior based on the data network geographical information (Specification page 15, line 21 – page 18, line 26; page 47, line 21 – page 48, line 22; Figure 6, all blocks).

E.   CLAIM 42 – INDEPENDENT

Claim 42 is directed to an apparatus for predicting customer behavior based on data network geographical influences. The apparatus includes (1) means for obtaining data network geographical information regarding a plurality of customers, the data network geographic information comprising frequency distributions of both (i) number of data network links between a customer geographical location and one or more web site data network geographical locations, and (ii) size of a click stream for arriving at the one or more web site data network geographical locations; (2) means for training a predictive algorithm using the data network geographical information; and (3) means for using the predictive algorithm to predict customer behavior based on the data network geographical information (Specification page 15, line 21 – page 18, line 26; page 47, line 21 – page 48, line 22; Figure 6, all blocks).

The structure corresponding to the means for obtaining is described at Specification page 45, lines 10-21 and depicted at 520 in Figure 5A. The structure corresponding to the means for training is described at Specification page 45, line 22 – page 47, line 13 and depicted at 530 and 540 in Figure 5A. The structure corresponding to the means for using is described at Specification page 47, lines 14-20 and depicted at 550 in Figure 5A.

F.   CLAIM 43 - INDEPENDENT

Claim 43 is directed to a computer program product in a computer readable medium. The computer program product includes instructions for enabling a data processing machine to predict customer behavior based on data network geographical influences, including: (1) instructions for obtaining data network geographical information regarding a plurality of customers, the data network geographic information comprising frequency distributions of both (i) number of data network links between a customer geographical location and one or more web site data network geographical locations, and (ii) size of a click stream for arriving at the one or more web site data network geographical locations; (2) instructions for training a predictive algorithm using the data network geographical information; and (3) instructions for using the predictive algorithm to predict customer behavior based on the data network geographical information (Specification page 15, line 21 – page 18, line 26; page 47, line 21 – page 48, line 22; Figure 6, all blocks).

GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The grounds of rejection to review on appeal are as follows:

1.   Whether Claims 1-8, 10-22, 24-35 and 37-43 were properly rejected under 35 U.S.C. § 101;

2.   Whether Claims 1, 15, 29 and 41-43 were properly rejected under 35 U.S.C. § 112, first paragraph;

3.   Whether Claims 1-8, 10-22, 24-35 and 37-40 were properly rejected as being obvious over *Menon et al.* (U.S. 5,537,488) in view of *Wu* (U.S. 6,741,967) and further in view of Appellant's background of the invention under 35 U.S.C. § 103(a); and

4.   Whether Claims 41-43 were properly rejected as being obvious over *Menon et al.* (U.S. 5,537,488) in view of Appellant's background of the invention and further in view of *Wu* (U.S. 6,741,967) under 35 U.S.C. § 103(a).

<u>**ARGUMENT**</u>

A. GROUND OF REJECTION 1 (Claims 1-8, 10-22, 24-35 and 37-43)

Claims 1-8, 10-22, 24-35 and 37-43 stand rejected under 35 U.S.C. § 101 as being directed
to non-statutory subject matter.

The PTO Interim Guidelines[1]

The Interim Guidelines establish a four step process for determining whether a claim
recites patentable subject matter: "1) Does the claimed invention fall within one of the four
statutory categories?[2] (i.e. process, machine, manufacture and composition of matter); 2) Does
the claimed invention fall within a judicial exception?[3] (i.e. law of nature, natural phenomena
and abstract idea); 3) If the invention is within a judicial exception, does it have a practical
application?[4] (i.e. does it result in a physical transformation or produce a **useful, concrete and
tangible result?**); and 4) If the invention is within a judicial exception, does it wholly preempt all
substantial applications of the judicial exception.[5] The test is applied in two stages. First the
examiner analyzes the claim to determine if it falls within one of the four statutory categories. If
not, the claim is directed to a non-statutory invention. If the claim is found to be a process,
machine, manufacture or composition of matter, the next stage of the test determines whether the
invention is within a judicial exception. For example, a claim to a mathematical algorithm may
be in the form of a process claim but it may also claim an abstract idea and, so, may not be
eligible for patent protection. Next, if the invention is covered by one of the judicial exceptions,
the examiner must determine if it has a practical application. This may be done in either of two

---

[1] 1300 Off. Gaz. Pat. Office 142 (Nov. 22, 2005)

[2] Interim Guidelines for Examination of Patent Applications for Patent Subject Matter Eligibility at § III
B

[3] Id. at § III C 1

[4] Id. at § III C 2

[5] Id. at § III C 3

ways, first by determining if the invention results in a physical transformation, for example, if the mathematical algorithm is used in a chemical process, and second if the invention produces a "useful, concrete and tangible result," (e.g., if the mathematical algorithm produces useful data concerning something physical and that data is then conveyed to a person). If the invention is found to be within one of the statutory categories and within one of the judicial exceptions but is also found to have a practical application, the analysis is not complete. The examiner must then ensure that the invention does not preempt every substantial practical application of the abstract idea, law of nature or natural phenomenon.

The Examiner erroneously characterizes all of claims 1-8, 10-22, 24-35 and 37-43 as being process claims that must somehow 'apply, involve, use, or advance the technological arts'. This 'not in the technological arts' criteria has been expressly superceded by the above described Interim Guidelines, as expressly stated in the USPTO's own overview of such guidelines, where Robert Weinhardt (USPTO Business Practice Specialist TC3600) expressly stated in such Overview[6]:

> "The following tests previously applied by some examiners are not determinative of patent –eligible subject matter and should not be used as rationale for rejecting claims under 35 USC 101.
>
> The "Not in the technological arts" criterion. <u>USPTO personnel should no longer rely on the technological arts test to determine whether a claimed invention is directed to statutory subject matter.</u> There are no other recognized exceptions to eligible subject matter other than laws of nature, natural phenomena, and abstract ideas" (emphasis added by Appellants)

As will be shown below, each of Claims 1-8, 10-22, 24-35 and 37-43 fully comply with the Interim Guidelines criteria for patentable subject matter.

---

[6] http://www.uspto.gov/web/offices/pac/compexam/interim_guide_subj_matter_eligibility.html

A.1. Claims 1-6, 8 and 10-14

With respect to Claims 1-6, 8 and 10-14, the Examiner states that such claims do not recite a concrete and tangible result, and that Claims 29 and 43 do not meet the definition of a true data structure (citing IEEE definition in MPEP 2106). Appellants urge error in such assertion, as described below.

For Claim 1, and per the Interim Guidelines test for patentable subject matter, a determination must be made as to "Does the claimed invention fall within one of the four statutory categories?". The answer is most definitely yes, as claim 1 recites "A data processing machine implemented method of selecting data sets for use with a predictive algorithm based on data network geographical information". Thus, Claim 1 is directed to a process, which is one of the four (4) explicitly enumerated classes of inventions spelled out in 35 U.S.C. 101. Next, per the Interim Guidelines, a determination must be made as to "Does the claimed invention fall within a judicial exception?[7] (i.e. law of nature, natural phenomena, and abstract idea)". Here, the answer quite simply is no, the claim does not fall within a judicial exception, as a data processing machine implemented method of selecting data sets for use with a predictive algorithm based on data network geographical information is neither a law of nature, a natural phenomena or an abstract idea. For example, Claim 1 is not merely directed to an abstract idea, as it expressly recites "A *data processing machine implemented method* of selecting data sets for use with a predictive algorithm based on data network geographical information" and thus is not a mere abstract idea. Because Claim 1 is not merely directed to one of the judicial exceptions of a law of nature, a natural phenomena or an abstract idea, the proper statutory subject matter analysis ends, and Claim 1 is statutory under 35 USC 101.

Even assuming arguendo that this Claim 1 is directed to an abstract idea (which Appellants deny), such claim in any event provides a practical application as it either provides a physical transformation or it produces a useful, concrete and tangible result and therefore provides a practical application, as will now be described in detail. Claim 1 recites *generating a first statistical distribution* of a training data set, which is a concrete and tangible result[8]. Claim

---

[7] Id. at § III C 1

[8] The term "tangible" is not limited to elements that may be perceived only by the sense of touch. To the contrary, the term "tangible" refers to anything that is capable of being perceived, precisely defined

1 also recites *generating a second statistical distribution* of a testing data set, which is also a concrete and tangible result. Claim 1 also recites *modifying selection of entries in one or more of the training data set and the testing data set based on the discrepancy between the first statistical distribution and the second statistical distribution*, such modified entries in the training/testing data set also being a concrete and tangible result, as well as a physical transformation of the data set(s), as the size of the data sets may be changed in response thereto (Specification page 24, middle paragraph).

Further yet, per MPEP 2106 and numerous judicial decisions, a machine claim is statutory when the machine, as claimed, produces a concrete, tangible and useful result (as in *State Street*, 149 F.3d at 1373, 47 USPQ2d at 1601) or when a specific machine is being claimed

---

or realized by the mind, or capable of being appraised at an actual or approximate value (see Merriam-Webster Online Dictionary Definition, a copy of which is included below). In other words, something is "tangible" if it is possible to verify its existence. This does not require that the element be "touchable" but merely "perceivable".

MERRIAM-WEBSTER ONLINE (www.Merriam-Webster.com) copyright 2005 by Merriam-Webster, Incorporated.

Main Entry: [1]**tan·gi·ble**
Pronunciation: 'tan-j&-b&l
Function: *adjective*
Etymology: Late Latin *tangibilis*, from Latin *tangere* to touch
1 a : capable of being perceived especially by the sense of touch : **PALPABLE** **b** : substantially real : **MATERIAL**
2 : capable of being precisely identified or realized by the mind <her grief was *tangible*>
3 : capable of being appraised at an actual or approximate value <*tangible* assets>
synonym see PERCEPTIBLE
- **tan·gi·bil·i·ty** /"tan-j&-'bi-l&-tE/ *noun*
- **tan·gi·ble·ness** /'tan-j&-b&l-n&s/ *noun*
- **tan·gi·bly** /-blE/ *adverb*

"concrete." *Dictionary.com Unabridged (v 1.1)*. Random House, Inc. 17 Jan. 2007.
<Dictionary.com http://dictionary.reference.com/browse/concrete>:

1. constituting an actual thing or instance; real: *a concrete proof of his sincerity.*
2. pertaining to or concerned with realities or actual instances rather than abstractions; particular (opposed to GENERAL): *concrete ideas.*
3. representing or applied to an actual substance or thing, as opposed to an abstract quality: *The words "cat," "water," and "teacher" are concrete, whereas the words "truth," "excellence," and "adulthood" are abstract.*

(as in *Alappat*, 33 F.3d at 1544, 31 USPQ2d at 1557). Claim 1 expressly recites "data processing *machine implemented method*", and therefore a specific machine is being claimed.

Thus, since Claim 1 does not fall within any of the judicial exceptions, and the claim explicitly recites a data processing machine implemented method (which is a process - one of the four definitions for statutory subject matter under 35 USC 101), Claim 1 is explicitly allowed per 35 USC 101 and does not fall into one of the three judicially determined exceptions. Even assuming arguendo that such claim is directed to an abstract idea (which it does not), such claim in any event provides both a physical transformation as well as a useful, concrete and tangible result, as described above. Accordingly, Claim 1 is statutory under 35 USC 101, and thus has been erroneously rejected under 35 USC 101.

A.2. Claim 7

In addition to the above reasons given above with respect to Claim 1 (of which Claim 7 depends upon, and such reasons are hereby incorporated by reference), Claim 7 recites additional concrete and tangible results, as it recites generating recommendations (a concrete and tangible result) for improving selection of entries in one or more of the training data set and the testing data set, and re-generating at least one of the first statistical distribution and the second statistical distribution based upon the recommendations (another concrete and tangible result). Thus, it is further urged that Claim 7 has been erroneously rejected under 35 USC 101 as it does in fact explicitly recite a machine implemented process that produces concrete and tangible results and thus has a practical application that does not wholly pre-empt an abstract idea.

A.3. Claims 15-22 and 24-28

With respect to Claim 15, such claim expressly recites an "apparatus for selecting data sets for use with a predictive algorithm based on data network geographical information", with the apparatus comprising a statistical engine and a comparison engine. An apparatus is a machine, which is expressly recognized by 35 USC 101 as being proper statutory subject matter[9]. This

---

[9] 35 U.S.C. 101: Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title (emphasis added by Appellants).

Claim 15 does not fall within a judicial exception, as it is not merely directed to a law of nature, a natural phenomena or an abstract idea, as it expressly recites "An apparatus for selecting data sets for use with a predictive algorithm based on data network geographical information, comprising: a statistical engine; a comparison engine coupled to the statistical engine". Accordingly, Claim 15 is statutory under 35 USC 101, and thus has been erroneously rejected under 35 USC 101.

A.4.   Claims 29-35 and 37-40

With respect to Claim 29, such claim expressly recites "A computer program product in a computer readable medium comprising instructions for enabling a data processing machine to select data sets for use with a predictive algorithm based on data network geographical information". It is urged that a claimed computer readable medium encoded with a computer program is a computer element which defines structural and functional inter-relationships between the computer program and the rest of the computer which permits the computer program's functionality to be realized, and is thus statutory. See *Lowry*, 32 F.3d at 1583-84, 32 USPQ2d at 1035, MPEP 2106, Interim Guidelines. Therefore, according to both Lowry and the MPEP, Claim 29 is statutory, and thus has been erroneously rejected under 35 USC 101.

A.5.   Claim 41

With respect to Claim 41, such claim recites "A data processing machine implemented method of predicting customer behavior based on data network geographical influences", and thus recites a machine implemented process which is one of the four statutorily defined categories of proper subject matter and does not fall within one of the three judicial exceptions. In addition, such claim recites obtaining data network geographical information regarding a plurality of customers, the data network geographic information comprising frequency distributions of both (i) number of data network links between a customer geographical location and one or more web site data network geographical locations, and (ii) size of a click stream for arriving at the one or more web site data network geographical locations - which are concrete and tangible results. In addition, Claim 41 recites using the predictive algorithm to predict customer behavior based on the data network geographical information – which is also a concrete and

tangible result. Thus, even to the extent such claim were improperly interpreted to be merely directed to an abstract idea, such claim provides a practical application, and is thus further shown to be statutory under 35 USC 101.

Further yet, per MPEP 2106 and numerous judicial decisions, a machine claim is statutory when the machine, as claimed, produces a concrete, tangible and useful result (as in *State Street*, supra) or when a specific machine is being claimed (as in Alappat, supra). Claim 41 expressly recites "data processing machine implemented method", and therefore a specific machine is being claimed.

Thus, since Claim 41 does not fall within any of the judicial exceptions, and the claim explicitly recites a data processing machine implemented method (which is a process - one of the four definitions for statutory subject matter under 35 USC 101), Claim 41 is explicitly allowed per 35 USC 101 and does not fall into one of the three judicially determined exceptions. Accordingly, Claim 41 is statutory under 35 USC 101, and thus has been erroneously rejected under 35 USC 101.

A.6.   Claim 42

With respect to Claim 42, such claim expressly recites an "apparatus for predicting customer behavior based on data network geographical influences". An apparatus is a machine, which is expressly recognized by 35 USC 101 as being proper statutory subject matter, and thus claim does not fall into one of the three judicial exceptions as it is specifically directed to an apparatus. Accordingly, Claim 42 is statutory under 35 USC 101, and thus has been erroneously rejected under 35 USC 101.

A.7.   Claim 43

With respect to Claim 43, such claim expressly recites, "A computer program product in a computer readable medium comprising instructions for enabling a data processing machine to predict customer behavior based on data network geographical influences". It is urged that a claimed computer readable medium encoded with a computer program is a computer element which defines structural and functional inter-relationships between the computer program and the rest of the computer which permits the computer program's functionality to be realized, and is thus statutory. See *Lowry*, 32 F.3d at 1583-84, 32 USPQ2d at 1035, MPEP 2106, Interim

Guidelines. Therefore, according to both Lowry and the MPEP, Claim 43 is statutory, and thus has been erroneously rejected under 35 USC 101.

B.   GROUND OF REJECTION 2 (Claims 1, 15, 29 and 41-43)

Claims 1, 15, 29 and 41-43 stand rejected under 35 U.S.C. § 112, first paragraph as failing to comply with the written description requirement.

While it is the Specification, and not the claims, that must comply with the written description requirement of 35 U.S.C. § 112, first paragraph, Appellants will in any event point out how the Specification does in fact comply with the written description requirement of 35 U.S.C. § 112, first paragraph.

B.1.   Claims 1, 15 and 29

As to Claims 1, 15 and 29, the Examiner state that nowhere in the Specification is it explained how the predictive algorithm would predict customer behavior based upon network geographic location. Appellants urge that this is not the case, as will now described in detail.

As shown in Figure 4, a set of customers 400 for which information has been obtained are present in a data network geographical area. These customers 400 are geographically located in the data network in clusters due to their affiliation with other customers that navigate the data network in a similar manner or are geographically located in the data network near other customers. For example, customers that navigate the data network using similar type search terms may be required to traverse the same number, or close to the same number, of links in order to arrive at a destination web site or web page. Because of this, these customers may be geographically located close to one another in the data network since it requires the same amount of travel distance for these customers to arrive at other data network web sites. From these customers 400 a customer database is generated 410 (Specification page 19, line 19 – page 20, line 8). From the customer database 410, a set of training data 420 and testing data 430 are generated. In known systems, these sets of data 420 and 430 are generated using a random selection process. Based on this random selection process, various ones of the customers in the customer database 410 are selected for inclusion into the training data set 420 and the testing data set 430. As can be seen from Figure 4, by selecting customers randomly from the customer database 410, it is possible that some of the clusters of customers may not be represented in the

training and testing data sets 420 and 430. Moreover, the training data set 420 and the testing data set 430 may not be commonly representative of the same clusters of customers. In other words, the training data set 420 may contain customers from clusters 1 and 3 while the testing data set 430 may contain customers selected from clusters 1 and 2. Because of the discrepancies between the training and testing data sets 420 and 430 with the customer database 410, certain types of customers may be over-represented and other types of customers may be under-represented. As a result, the predictive algorithm may not accurately represent the behavior of potential customers. Moreover, because of the discrepancies between the training and testing data sets 420 and 430, the predictive algorithm may be trained improperly. That is, the training data set 420 may train the predictive algorithm to output a particular predicted customer behavior based on a particular input. However, the testing data set 430 may indicate a different customer behavior based on the same input due to the differences in the customer clusters represented in the training data set 420 and the testing data set 430 (Specification page 20, line 19 – page 21, line 27).

For example, as shown in Figure 4, the training data set 420 is predominantly comprised of customers from clusters A, B and C. Assume that customers in clusters A and B are very good customer candidates for new electronic items while customers in group C are only mildly good customer candidates for new electronic items. Based on this training data, if a commercial web site at data network location X were interested in introducing a new electronic item, the predictive algorithm may indicate that there is a high likelihood of customer demand for the new electronic item from customers in clusters A and B. However, in actuality, assume that customers in clusters A and B are less likely to navigate the data network from their data network location to the data network location X due to the amount of interaction required, i.e. the size of the user click stream. Thus, the predictive algorithm will provide an erroneous result. Moreover, if the testing data contains customers from clusters A, B, D and E, the customer behaviors in the testing data will be different from that of customers in the training data set (comprising clusters A, B and C). As a result, the testing data and the training data are not consistent and erroneous customer behavior predictions will arise. Thus, data network geographic effects of clustering must be taken into account *when selecting customers to be included* in training and testing data sets of a customer behavior predictive algorithm (Specification page 22, line 1 – page 23, line 3). With the present invention, the discrepancies between a testing data set and a training data set are

identified. Furthermore, the discrepancies between both the testing data set and the training data set and the customer database are identified. *The discrepancies are identified based on a data network geographical characteristic such as a number of links or the size of a user click stream.* The normalized frequency distributions of the number of links and/or user click stream in the training data set are compared to the normalized frequency distributions of the testing data set. If the differences between the frequency distributions are above a predetermined tolerance, the two data sets are too different to provide accurate training of the predictive algorithm *when taking data network geographical influences into account.* This same procedure may be performed with regard to the frequency distribution of the customer database (Specification page 23, lines 4 – 21).

In order to compare the frequency distributions, the mean, mode and/or standard deviations of the frequency distributions may be compared with one another to determine if the frequency distributions are similar within a predetermined tolerance. The mean is a representation of the average of the frequency distribution. The mode is a representation of the most frequently occurring value in the data set. The standard deviation is a measure of dispersion in a set of data. Based on these quantities for each frequency distribution, a comparison of the frequency distributions may be made to determine if they adequately represent the customer population clusters in the customer database. *If they do not, the present invention may, based on the relative discrepancies of the various data sets, make recommendations as to how to better select training and testing data sets that represent the data network geographic clustering of customers.* For example, if the relative discrepancy between a testing data set and a training data set are such that the training data set does not contain enough customers to represent all of the necessary clusters in the testing data set, the training data set may need to be increased in size. Similarly, if the testing data set and/or training data set do not contain enough customers to represent all of the clusters in the customer database, the testing and training data sets may need to be increased. In such cases, the same random selection algorithm may be used and the same seed value of the random selection algorithm may be used with the number of customers selected being increased. Moreover, the testing data set and training data sets may be combined to form a composite data set, which may be compared to the customer database. In combining the two data sets, customers appearing in a first data set, and not in the second data set, are added to the composite data set, and vice versa (Specification page 23, line 23 – page 25. line 4).

The frequency distribution of the composite data set may be compared to the frequency distribution of the customer data set, in the manner described above, to determine if the composite represents the customer clusters appropriately. If the composite data set does represent the customer clusters of the customer database appropriately, the composite data set may be used to train the predictive algorithm. If the composite data set does not represent the customer clusters of the customer database appropriately, a new random selection algorithm may need to be used or a new seed value of a random selection algorithm may need to be used. *In this way, the selection of training and testing data is modified such that the training and testing data better represents actual customer behavior based on data network geographical influences* (Specification page 25, lines 5-20).

Figure 6 is a flowchart outlining an exemplary operation of the present invention. As shown in Figure 6, the operation starts with gathering customer database information (step 610). The customer database information is then used as a basis for selecting a training data set and/or testing data set (step 620). Frequency distribution information of a number of data network links and/or user click stream to a web site of interest is calculated for each of the training data set, testing data set and customer database data set (step 630). The frequency distribution information for each of these data sets is compared and evaluated to determine if differences exceed a predetermined tolerance (step 640). *A determination is made as to whether differences in the frequency distribution information is beyond a predetermined tolerance (step 650). If so, recommendations are generated based on the particular differences (step 660) and the operation returns to step 620 where the training and testing data sets are again determined in view of the recommendations.* If the differences in frequency distribution information are not beyond the predetermined tolerance, the training data set and testing data set are used to train the predictive algorithm (step 670) and the operation ends. Thereafter, the predictive algorithm may be used to generate customer behavior predictions *taking into account the data network geographical influences of customers as represented in the training and testing data sets* (page 47, line 21 – page 48, line 22).

Therefore, the rejection of Claims 1, 15 and 29 under 35 U.S.C. § 112, first paragraph is clearly erroneous, as the Specification does in fact describe in detail *how parameters used by the predictive algorithm – and in particular the training data sets and testing datasets that are used to train the predictive algorithm - are modified to improve predicted customer behavior based upon network geographic location.* Once the predictive algorithm has been trained, by taking into account the data network geographical influences of customers as represented in the training and testing data sets, it is then used in a standard, normal fashion as is commonly known to those in the data mining art (Specification page 47, lines 14-20). Restated, it is the training of the predictive algorithm using data network geographic information that is a key aspect of the inventive concepts described in the Specification and claimed in the claims. The subsequent use of such predictive algorithm - after being uniquely trained per the claimed features recited in the present application - is a standard use of a data mining predictive algorithm, and which is commonly known to those of ordinary skill in the art[10]. Thus, it is urged that Claim 1 has been erroneously rejected under 35 U.S.C. § 112, first paragraph.

B.2. Claims 41-43

As to Claims 41-43, the Examiner states that nowhere in the Specification is it explained how the predictive algorithm would predict customer behavior based upon network geographic location. Appellants urge that this is not the case, as will now described in detail.

As shown in Figure 4, a set of customers 400 for which information has been obtained are present in a data network geographical area. These customers 400 are geographically located in the data network in clusters due to their affiliation with other customers that navigate the data network in a similar manner or are geographically located in the data network near other customers. For example, customers that navigate the data network using similar type search terms may be required to traverse the same number, or close to the same number, of links in order to arrive at a destination web site or web page. Because of this, these customers may be

---

[10] The law is clear that patent documents need not include subject matter that is known in the field of the invention and is in the prior art, for patents are written for persons experienced in the field of the invention. *Vivid Technologies, Inc. v. American Science and Engineering, Inc.*, 200 F.3d 795, 804, 53 USPQ2d 1289, 1295 (Fed. Cir. 1999). The enablement requirement ensures ... that a specification shall disclose an invention in such a manner as will enable one skilled in the art to make and utilize it. *In re Gay*, 309 F.2d 769, 772, 135 USPQ 311, 315 (CCPA 1962); *Spectra-Physics,Inc. v. Coherent, Inc.*, 827

geographically located close to one another in the data network since it requires the same amount of travel distance for these customers to arrive at other data network web sites. From these customers 400 a customer database is generated 410 (Specification page 19, line 19 – page 20, line 8). From the customer database 410, a set of training data 420 and testing data 430 are generated. In known systems, these sets of data 420 and 430 are generated using a random selection process. Based on this random selection process, various ones of the customers in the customer database 410 are selected for inclusion into the training data set 420 and the testing data set 430. As can be seen from Figure 4, by selecting customers randomly from the customer database 410, it is possible that some of the clusters of customers may not be represented in the training and testing data sets 420 and 430. Moreover, the training data set 420 and the testing data set 430 may not be commonly representative of the same clusters of customers. In other words, the training data set 420 may contain customers from clusters 1 and 3 while the testing data set 430 may contain customers selected from clusters 1 and 2. Because of the discrepancies between the training and testing data sets 420 and 430 with the customer database 410, certain types of customers may be over-represented and other types of customers may be under-represented. As a result, the predictive algorithm may not accurately represent the behavior of potential customers. Moreover, because of the discrepancies between the training and testing data sets 420 and 430, the predictive algorithm may be trained improperly. That is, the training data set 420 may train the predictive algorithm to output a particular predicted customer behavior based on a particular input. However, the testing data set 430 may indicate a different customer behavior based on the same input due to the differences in the customer clusters represented in the training data set 420 and the testing data set 430 (Specification page 20, line 19 – page 21, line 27).

For example, as shown in Figure 4, the training data set 420 is predominately comprised of customers from clusters A, B and C. Assume that customers in clusters A and B are very good customer candidates for new electronic items while customers in group C are only mildly good customer candidates for new electronic items. Based on this training data, if a commercial web site at data network location X were interested in introducing a new electronic item, the predictive algorithm may indicate that there is a high likelihood of customer demand for the new

---

F2d 1524, 1532, 3 USPQ2d 1737, 1742 (Fed. Cir. 1987).

electronic item from customers in clusters A and B. However, in actuality, assume that customers in clusters A and B are less likely to navigate the data network from their data network location to the data network location X due to the amount of interaction required, i.e. the size of the user click stream. Thus, the predictive algorithm will provide an erroneous result. Moreover, if the testing data contains customers from clusters A, B, D and E, the customer behaviors in the testing data will be different from that of customers in the training data set (comprising clusters A, B and C). As a result, the testing data and the training data are not consistent and erroneous customer behavior predictions will arise. Thus, data network geographic effects of clustering must be taken into account *when selecting customers to be included* in training and testing data sets of a customer behavior predictive algorithm (Specification page 22, line 1 – page 23, line 3). With the present invention, the discrepancies between a testing data set and a training data set are identified. Furthermore, the discrepancies between both the testing data set and the training data set and the customer database are identified. *The discrepancies are identified based on a data network geographical characteristic such as a number of links or the size of a user click stream.* The normalized frequency distributions of the number of links and/or user click stream in the training data set are compared to the normalized frequency distributions of the testing data set. If the differences between the frequency distributions are above a predetermined tolerance, the two data sets are too different to provide accurate training of the predictive algorithm *when taking data network geographical influences into account.* This same procedure may be performed with regard to the frequency distribution of the customer database (Specification page 23, lines 4 – 21).

In order to compare the frequency distributions, the mean, mode and/or standard deviations of the frequency distributions may be compared with one another to determine if the frequency distributions are similar within a predetermined tolerance. The mean is a representation of the average of the frequency distribution. The mode is a representation of the most frequently occurring value in the data set. The standard deviation is a measure of dispersion in a set of data. Based on these quantities for each frequency distribution, a comparison of the frequency distributions may be made to determine if they adequately represent the customer population clusters in the customer database. *If they do not, the present invention may, based on the relative discrepancies of the various data sets, make recommendations as to how to better select training and testing data sets that represent the data network geographic*

clustering of customers. For example, if the relative discrepancy between a testing data set and a training data set are such that the training data set does not contain enough customers to represent all of the necessary clusters in the testing data set, the training data set may need to be increased in size. Similarly, if the testing data set and/or training data set do not contain enough customers to represent all of the clusters in the customer database, the testing and training data sets may need to be increased. In such cases, the same random selection algorithm may be used and the same seed value of the random selection algorithm may be used with the number of customers selected being increased. Moreover, the testing data set and training data sets may be combined to form a composite data set which may be compared to the customer database. In combining the two data sets, customers appearing in a first data set, and not in the second data set, are added to the composite data set, and vice versa (Specification page 23, line 23 – page 25. line 4).

The frequency distribution of the composite data set may be compared to the frequency distribution of the customer database, in the manner described above, to determine if the composite represents the customer clusters appropriately. If the composite data set does represent the customer clusters of the customer database appropriately, the composite data set may be used to train the predictive algorithm. If the composite data set does not represent the customer clusters of the customer database appropriately, a new random selection algorithm may need to be used or a new seed value of a random selection algorithm may need to be used. *In this way, the selection of training and testing data is modified such that the training and testing data better represents actual customer behavior based on data network geographical influences* (Specification page 25, lines 5-20).

Figure 6 is a flowchart outlining an exemplary operation of the present invention. As shown in Figure 6, the operation starts with gathering customer database information (step 610). The customer database information is then used as a basis for selecting a training data set and/or testing data set (step 620). Frequency distribution information of a number of data network links and/or user click stream to a web site of interest is calculated for each of the training data set, testing data set and customer database data set (step 630). The frequency distribution information for each of these data sets is compared and evaluated to determine if differences exceed a predetermined tolerance (step 640). *A determination is made as to whether differences in the frequency distribution information is beyond a predetermined tolerance (step 650). If so, recommendations are generated based on the particular differences (step 660) and the operation*

returns to step 620 where the training and testing data sets are again determined in view of the recommendations. If the differences in frequency distribution information are not beyond the predetermined tolerance, the training data set and testing data set are used to train the predictive algorithm (step 670) and the operation ends. Thereafter, the predictive algorithm may be used to generate customer behavior predictions *taking into account the data network geographical influences of customers as represented in the training and testing data sets* (page 47, line 21 – page 48, line 22).

Therefore, the rejection of Claims 41-43 under 35 U.S.C. § 112, first paragraph is clearly erroneous, as the Specification does in fact describe in detail *how parameters used by the predictive algorithm – and in particular the training data sets and testing datasets that are used to train the predictive algorithm - are modified to improve predicted customer behavior based upon network geographic location.*

C.   GROUND OF REJECTION 3 (Claims 1-8, 10-22, 24-35 and 37-40)

Claims 1-8, 10-22, 24-35 and 37-40 stand rejected under 35 U.S.C. § 103(a) as being obvious over *Menon et al.* (U.S. 5,537,488) in view of *Wu* (U.S. 6,741,967) and further in view of Appellant's background of the invention.

C.1.   Claims 1, 12-15, 26-29, 39 and 40

The present invention of Claim 1 is directed to an improved technique for selecting data sets for use with a predictive algorithm. A *statistical distribution of a training data set* is compared with a *statistical distribution of a testing data set* to identify a discrepancy between these distributions with respect to data network geographic information. *Based upon such comparison and its associated discrepancy identification, selection of entries in the training data set and/or testing data set are modified.* These modified entries are then used by the predictive algorithm, thereby taking into account the influences of data network geography when using the predictive algorithm. None of the cited references makes any mention of *using data network geographic information to modify entries* of the testing or training data sets that are used by a predictive algorithm.

Specifically, Claim 1 recites, "using the first statistical distribution and the second statistical distribution to identify a discrepancy between the first statistical distribution and the

second statistical distribution with respect to the data network geographical information". The Examiner states that *Menon* teaches "using the first statistical distribution and the second statistical distribution to identify a discrepancy between the first statistical distribution and the second statistical distribution" at column 20, lines 61-64. Appellants urge twofold error in such assertion. *First*, this cited passage does not teach or suggest two different statistical distributions, and Claim 1 expressly recites using *both* the first statistical distribution (of the training data set) *and* the second statistical distribution (of the testing data set). This cited *Menon* passage describes receiving one test input pattern (which is not a *statistical distribution* of a testing data set, as claimed) and computing a correlation between (i) this input test pattern and (ii) a category definition. The *Menon* category definition is not a *statistical distribution* of a training data set, as claimed. Thus, this cited passage does not teach any use of two statistical distributions. *Second,* even if the above assertion were true, Claim 1 goes further and recites that the identified discrepancy between these two statistical distributions is *with respect to the data network geographic information.* The Examiner acknowledges that the cited *Menon* reference does not teach data network geographic information, but states that the cited *Wu* reference teaches data network geographic information. Appellants urge that even if true, the existence of data network geographic information as per the teachings of *Wu* does not teach or suggest the synergistic co-action between the claimed (1) first statistical distribution of a training data set, (2) second statistical distribution of a testing data set, and (3) data network geographic information. Instead, the resulting combination teaches computing a correlation between a category definition and a single test input pattern, where the category/test pattern pertains to data network geographic information. Such resulting combination does not teach or suggest, "using the first statistical distribution and the second statistical distribution *to identify a discrepancy between the first statistical distribution and the second statistical distribution with respect to the data network geographical information*". It is therefore respectfully submitted that the Examiner has failed to properly establish a prima facie showing of obviousness with respect to Claim 1[11]. Accordingly,

---

[11] To establish prima facie obviousness of a claimed invention, all of the claim limitations must be taught or suggested by the prior art (emphasis added by Appellants). MPEP 2143.03. *See also, In re Royka,* 490 F.2d 580 (C.C.P.A. 1974).

the burden has not shifted to Appellants to overcome an obviousness assertion[12]. In addition, as a proper prima facie showing of obviousness has not been established, Claim 1 has been improperly rejected[13].

Still further, the details of how this using step (with respect to the first statistical distribution and the second statistical distribution) is accomplished are substantially different from what is taught by the cited references. Claim 1 expressly recites "using the first statistical distribution and the second statistical distribution *to identify a discrepancy* between the first statistical distribution and the second statistical distribution with respect to the data network geographical information *by comparing* at least one of the first statistical distribution and the second statistical distribution to a statistical distribution of a customer database to determine if at least one of the training data set and the testing data set are geographically representative of a customer population represented by the customer database". As can be seen, as a part of the discrepancy identification, at least one of the first statistical distribution and the second statistical distribution is compared to a statistical distribution of a customer database to determine if at least one of the training data set and the testing data set are geographically representative of a customer population represented by the customer database. In rejecting this comparing aspect of Claim 1, the Examiner states:

"using the modified selection of entries by the predictive algorithm and that said using is done by comparing by comparing at least one of the first statistical distribution and the second statistical distribution to a statistical distribution of a customer database"

As can be seen, the alleged 'comparing' step is with respect to the predictive algorithm's use of a modified selection of entries. In contrast, the claimed 'comparing' step is with respect to discrepancy determination between the first statistical distribution and the second statistical

---

[12] In rejecting claims under 35 U.S.C. Section 103, the examiner bears the initial burden of presenting a prima facie case of obviousness. *In re Oetiker*, 977 F.2d 1443, 1445, 24 USPQ2d 1443, 1444 (Fed. Cir. 1992). Only if that burden is met, does the burden of coming forward with evidence or argument shift to the Appellant. *Id.*

[13] If the examiner fails to establish a prima facie case, the rejection is improper and will be overturned.

distribution with respect to the data network geographical information (which is a different claimed step that is in addition to the predictive algorithm using step). Thus, it is further urged that the Examiner has failed to properly establish a prima facie showing of obviousness, as the 'comparing' step is alleged to be with respect to 'using' of a predictive algorithm, whereas what is actually claimed is that the comparing step is with respect 'using' a first a first statistical distribution and a second statistical distribution to identify a discrepancy between the first statistical distribution and the second statistical distribution with respect to the data network geographical information. The Examiner has not even alleged such a teaching or suggestion.

Further yet, Claim 1 recites, "modifying selection of entries in one or more of the training data set and the testing data set based on the discrepancy between the first statistical distribution and the second statistical distribution". In rejecting this aspect of Claim 1, the Examiner cites *Menon's* teaching at col. 21, lines 20-24 as teaching this claimed selection of entries modification step. Appellants respectfully submit that this passage states that a new category is defined. A category is not a training data set or a testing data set. While it is true that input training patterns are received and grouped into clusters, and each cluster is associated with a category (col. 1, lines 24-28), these categories are not used by a predictive algorithm. In contrast, per the features of Claim 1, the modified entries of the testing or training data set are used by the predictive algorithm, thereby advantageously improving the predictive algorithms ability to predict based on data network geographic information. Quite simply, the definition of a new category as described by *Menon* does not teach or suggest *modifying selection of entries in a training or testing data set which is then used by a predictive algorithm*.

Further with respect to Claim 1, Appellants urge that none of the cited references teach (or otherwise suggest) the claimed step of "comparing at least one of the first statistical distribution and the second distribution to a distribution of a customer database to determine if at least one of the training data set and the testing data set *are geographically representative of a customer population* represented by the customer database". As can be seen, this claimed feature is directed to comparing one or more of the first and second distributions (of the testing and training data sets) *with another distribution* – the distribution of a customer database in order to determine if there is a proper geographic representation of the customer population represented

---

In re Fine, 837 F.2d 1071, 1074, 5 USPQ2d 1596, 1598 (Fed. Cir. 1988).

by the customer database. The cited *Menon* reference does not teach (or otherwise suggest) a *distribution* of a customer database, and hence it necessarily follows that it does not teach (or otherwise suggest) any comparing step being made with such (missing) distribution of a customer database. In rejecting Claim 1, the Examiner cites *Menon* col. 5, line 35 – col. 6, line 56 and col. 6, line 57 – column 7, line 21 as teaching these features of Claim 1. Appellants urge that such passages describe details of how to group training patterns into categories in order to generate a training histogram, as claimed by *Menon* in Claim 24 (col. 20, lines 55-60). These cited passages deal with training patterns and the labeling of these training patterns' associated categories. The calculations described by *Menon* are *only with respect to training patterns* – albeit organized into different groups or categories. Importantly, there is no teaching (or suggestion) of *comparing such training patterns to a **distribution** of a customer database*, as expressly recited in Claim 1. Thus, it is further shown that a prima facie case of obviousness has not been established with respect to Claim 1.

Thus, a proper prima facie showing of obviousness has not been established by the Examiner, for the numerous reasons articulated above, and accordingly Claim 1 has been erroneously rejected.

Still further, Appellants urge that the Examiner is using improper hindsight analysis in rejecting Claim 1. The cited *Menon* reference is directed to techniques for pattern recognition for a person recognition system. A person of ordinary skill in the art, when presented with such pattern recognition techniques, would not have been motivated to somehow selectively transform and further modify such a system in order to adopt such teachings for use in predicting customer behavior based on network characteristics. It is error to reconstruct the patentee's claimed invention from the prior art by using the patentee's claims as a "blueprint". When prior art references require selective combination to render obvious a subsequent invention, there must be some reason for the combination other than the hindsight obtained from the invention itself. *Interconnect Planning Corp. v. Feil*, 774 F.2d 1132, 227 USPQ 543 (Fed. Cir. 1985). The fact that a prior art device could be modified so as to produce the claimed device is not a basis for an obviousness rejection unless the prior art suggested the desirability of such a modification. *In re Gordon*, 733 F.2d 900, 221 USPQ 1125 (Fed. Cir. 1984). Although a device may be capable of being modified to run the way [the patent Appellant's] apparatus is claimed, there must be a

suggestion or motivation *in the reference* to do so. *In re Mills*, 916 F.2d 680, 16 USPQ2d 1430 (Fed. Cir. 1990). The only reason for such modification – in effect combining two unrelated references which are directed to completely different systems (one being a person recognition system; the other being a system for designing web site test cases) - is coming from Appellant's own disclosure, which is impermissible hindsight analysis.

The Examiner themselves use Appellant's own disclosure in the background section of the present patent application as the catalyst for making the combination of such dissimilar teachings – further evidencing improper hindsight analysis (Office Action dated 09/06/2006, bottom of page 5 extending to the top of page 6). Quite simply, a person of ordinary skill in the person recognition art would not have been motivated to include teachings from a web site test case generation technique as such teachings are not related to one another without the benefit of Appellant's own disclosure as the catalyst to make such an unnatural combination. Thus, it is further urged that Claim 1 has been erroneously rejected using impermissible hindsight analysis.

Even when impermissibly using Appellant's own disclosure as the catalyst for the obviousness rejection, there are still missing claimed features not taught or suggested by the cited references – strongly evidencing non-obviousness. For example, in the most recent Office Action dated June 1, 2007, which was a result of an Appeal Brief filed by Appellants pointing out the numerous differences between the features of Claim 1 and the cited references, the Examiner opines in their rebuttal to such Appeal Brief arguments:

> "The Applicant argues that none of the references make any mention of using data network geographic information to modify entries of the testing or training data sets that are used by a predictive algorithm. The Examiner answers that Applicant's <u>background of the Invention</u> teaches that it is old and well known in the artificial intelligence art to input training and test data into a predictive algorithm for the purpose of predicting a customer's propensity to respond to an offer or his propensity to buy a product (see Applicant's <u>background</u> page 3)"

Appellants urge that they are not merely claimed the use of training and test data by a predictive algorithm, but instead the features of Claim 1 further enhance such techniques by *modifying selection of entries* in one or more of the training data set and the testing data set *based on the*

discrepancy between the first statistical distribution and the second statistical distribution.
Thus, the Examiner's reliance on Appellant's background of the Invention to establish what was old and well-known fails to overcome the teaching/suggestion deficiencies identified hereinabove. Quite simply, it is the particular usage (modifying selection of testing/training data set entries) of the particular data (discrepancy between the first statistical distribution of a training data set and the second statistical distribution of a testing data set with respect to the data network geographical information) that is not taught or suggested by the combined teachings of the cited references. Thus, Claim 1 has been erroneously rejected.

C.2.   Claims 2, 16 and 30

In addition to the reasons given above with respect to Claim 1 (of which Claim 2 depends upon), Appellants further urge error in the rejection of Claim 2, as such claim recites "wherein the first statistical distribution and the second statistical distribution are distributions of a number of data network links from a customer data network geographical location to a web site data network geographical location". In rejecting Claim 2, the Examiner states that _Wu_ teaches a system that determines customer's navigational path through websites or web pages by calculating the amount of links by task, site and speed of search results in order to predict if an increase in a customer's purchase rate was a result of an improvement is the navigational path, citing _Wu's_ teaching at column 18, table B and column 36, lines 24-30. Even assuming arguendo that such assertion is true, this still does not establish any teaching or suggestion of the specific claimed feature that the first and second _statistical distributions that are used to compare to a statistical distribution of a customer_ database _to determine if at least one of the training data set and the testing data set are geographically representative of a customer population represented by the customer database_ are themselves 'distributions of a number of data network links from a customer data network geographical location to a web site data network geographical location'. Simply put, even if _Wu_ is alleged to teach a calculation of the amount of links by task, such alleged teaching does not establish a teaching/suggestion of the _specific use_ of such link information, as expressly recited in Claim 2. Thus, a proper prima facie showing of obviousness has not been established by the Examiner, and accordingly Claim 2 has been erroneously rejected.

C.3. Claims 3, 17 and 31

In addition to the reasons given above with respect to Claim 1 (of which Claim 3 depends upon), Appellants further urge error in the rejection of Claim 3, as such claim recites "wherein the first statistical distribution and the second statistical distribution are distributions of a size of a click stream for arriving at a web site data network geographical location". In rejecting Claim 3, the Examiner states that *Wu* teaches a system that determines customer's navigational path through websites or web pages by calculating the amount of links by task, site and speed of search results in order to predict if an increase in a customer's purchase rate was a result of an improvement is the navigational path, citing *Wu's* teaching at column 18, table B and column 36, lines 24-30. Even assuming arguendo that such assertion is true, this still does not establish any teaching or suggestion of the specific claimed feature that the first and second *statistical distributions that are used to compare to a statistical distribution of a customer* database *to determine if at least one of the training data set and the testing data set are geographically representative of a customer population represented by the customer database* are themselves 'distributions of a size of a click stream for arriving at a web site data network geographical location'. Simply put, even if *Wu* is alleged to teach a calculation of the size of a click stream, such teaching does not establish a teaching/suggestion of the *specific use* of such click stream information, as expressly recited in Claim 3. Thus, a proper prima facie showing of obviousness has not been established by the Examiner, and accordingly Claim 3 has been erroneously rejected.

C.4. Claims 4, 18 and 32

In addition to the reasons given above with respect to Claim 1 (of which Claim 4 depends upon), Appellants further urge error in the rejection of Claim 4, as such claim recites "wherein comparing the first statistical distribution and the second statistical distribution includes comparing one or more of a mean, mode, and standard deviation of the first statistical distribution to one or more of a mean, mode, and standard deviation of the second statistical distribution". As can be seen, Claim 4 further refines the comparing step recited in Claim 1, *such comparing step being between two statistical distributions* (a first statistical distribution of a training data set and a second statistical distribution of a testing data set). In rejecting Claim 4, the Examiner cites *Menon's* teaching at column 6, line 57 – column 7, line 20). Appellants

respectfully urge that while this cited passage mentions a 'mean', this passage teaches *normalization* for training data sets only (there is no mention of testing data sets, nor is there any mention of *comparing statistical distributions of both training data sets and testing data sets*). Quite simply, this passage teaches use of a mean training data set in an unrelated activity (normalization of such training data set). Thus, a proper prima facie showing of obviousness has not been established by the Examiner, and accordingly Claim 4 has been erroneously rejected.

C.5. Claims 5, 19 and 33

In addition to the reasons given above with respect to Claim 1 (of which Claim 5 depends upon), Appellants further urge error in the rejection of Claim 5, as such claim recites "wherein the first statistical distribution and the second statistical distribution are distributions of a weighted data network geographical distance between a customer data network geographical location and a web site data network geographical locations". In rejecting Claim 5, the Examiner states that *Wu* teaches a system that determines customer's navigational path through websites or web pages by calculating the amount of links by task, site and speed of search results in order to predict if an increase in a customer's purchase rate was a result of an improvement is the navigational path, citing *Wu's* teaching at column 18, table B and column 36, lines 24-30. Appellants respectfully urge that such link calculation allegation does not address the specific claimed feature recited in Claim 5 pertaining to a *weighted data network geographical distance* between a customer data network geographical location and a web site data network geographical locations. Thus, a proper prima facie showing of obviousness has not been established with respect to the claimed weighted distance feature recited in Claim 5.

Still further, and even assuming arguendo that the cited reference teaches a weighted distance feature (which it does not), this still does not establish any teaching or suggestion of the specific claimed feature that the first and second *statistical distributions that are used to compare to a statistical distribution of a customer* database *to determine if at least one of the training data set and the testing data set are geographically representative of a customer population represented by the customer database* are themselves 'distributions of a weighted data network geographical distance between a customer data network geographical location and a web site data network geographical locations'. Simply put, even if *Wu* did teach a weighted distance feature (which it does not), the existence of such feature does not establish a

teaching/suggestion of the *specific _use_* of such weighted distance, as expressly recited in Claim 5. Thus, a proper prima facie showing of obviousness has not been established by the Examiner, and accordingly Claim 5 has been erroneously rejected.

C.6.   Claims 6, 20 and 34

In addition to the reasons given above with respect to Claim 1 (of which Claim 6 depends upon), Appellants further urge error in the rejection of Claim 6, as such claim recites "wherein the first statistical distribution and the second statistical distribution are distributions of a weighted click stream for arriving at a web site data network geographical locations". In rejecting Claim 6, the Examiner states that *Wu* teaches a system that determines customer's navigational path through websites or web pages by calculating the amount of links by task, site and speed of search results in order to predict if an increase in a customer's purchase rate was a result of an improvement is the navigational path, citing *Wu's* teaching at column 18, table B and column 36, lines 24-30. Appellants respectfully urge that such link calculation allegation does not address the specific claimed feature recited in Claim 6 pertaining to a *weighted click stream*. Thus, a proper prima facie showing of obviousness has not been established with respect to the claimed weighted click stream feature recited in Claim 6.

Still further, and even assuming arguendo that the cited reference teaches a weighted click stream feature (which it does not), this still does not establish any teaching or suggestion of the specific claimed feature that the first and second *statistical distributions that are used to compare to a statistical distribution of a customer* database *to determine if at least one of the training data set and the testing data set are geographically representative of a customer population represented by the customer database* are themselves 'distributions of a weighted click stream for arriving at a web site data network geographical locations'. Simply put, even if *Wu* did teach a weighted click stream feature (which it does not), the existence of such feature does not establish a teaching/suggestion of the *specific _use_* of such weighted click stream, as expressly recited in Claim 6. Thus, a proper prima facie showing of obviousness has not been established by the Examiner, and accordingly Claim 6 has been erroneously rejected.

C.7. Claims 7, 21 and 35

In addition to the reasons given above with respect to Claim 1 (of which Claim 7 depends upon), Appellants further urge error in the rejection of Claim 7, as none of the cited references teach or suggest the claimed feature of *generating recommendations for improving selection of entries* in either or both of the training and testing data set. The passage cited by the Examiner in rejecting Claim 7 merely states that a new category is defined (if the correlation is below a threshold). Such 'definition of a new category' does not provide any type of *recommendation for improving selection of entries*, as expressly recited in Claim 7, and thus Claim 7 is further shown to have been erroneously rejected as there are additional claimed features not taught or suggested by any of the cited references.

Still further, such definition of a new category does not teach or otherwise suggest the claimed feature of "re-generating at least one of the first statistical distribution and the second statistical distribution based upon the recommendations". As can be seen, a *statistical distribution is re-generated* based upon the recommendation. Since there is no teaching of a recommendation, there is no teaching of performing an action (re-generating a statistical distribution) based upon such (missing) recommendation. Further, the definition of a new category does not pertain in any way to re-generating a *statistical distribution* that is used to identify discrepancies between the first statistical distribution and the second statistical distribution with respect to the data network geographical information, as expressly required by Claim 7 in combination with Claim 1. Accordingly, Claim 7 has been erroneously rejected as a proper prima facie showing of obviousness has not been established by the Examiner.

C.8. Claims 8 and 22

In addition to the reasons given above with respect to Claim 1 (of which Claim 8 depends upon), Appellants further urge error in the rejection of Claim 8, as such claim recites "wherein the training data set and the testing data set are selected from a customer information database comprising information with respect to customers who have purchased any of of goods and services over a data network, wherein the data network geographic information pertains to geographic information of the data network". As can be seen, *both* the training data set and the testing data set are selected from a customer information database (where this customer information database contains information pertaining to goods/services purchased over a data network). In rejecting

the 'customer information database selection' aspect of this claim – where both the training data set and the testing data set are selected – the Examiner cites *Menon's* teaching at col. 5, lines 37-55 as teaching such selection. Appellants urge that there, *Menon* states:

> "When the system of the present invention is trained, it receives **training** data patterns from various subjects or classes. In the case of a face recognition system, these patterns may include photographs of individual persons from several different orientations and/or exhibiting several different facial expressions. Photographs may also be shown of subjects with and without eyeglasses, with and without facial hair, etc. Voice data from different persons (classes) can also be received. As another example, in the case of a system used to identify semiconductor wafer defects, visual images of different types of defects as well as images of wafers having no defects can be received for training.
>
> Each **training** pattern is associated with a known class and takes the form of a feature pattern vector I_{INP}. Each category definition I.sub.k is expressed in a vector format compatible with the feature vector. As each pattern vector is received, a correlation C_{TRN} between it and each existing category definition is performed. In the case of a face recognition system, the correlation is computed according to"

As can be seen, this passage merely describes actions associated with *training* data patterns. There is no mention of any type of *testing* data patterns, and thus this cited passage does not teach the claimed feature of "the training data set and the testing data set are selected from a customer information database" (emphasis added), as erroneously alleged by the Examiner to be taught by this cited *Menon* passage.

C.9. Claims 10, 24 and 37

In addition to the reasons given above with respect to Claim 1 (of which Claim 10 depends upon), Appellants further urge error in the rejection of Claim 10, as such claim recites "wherein the first statistical distribution and second statistical distribution are frequency distributions of number of data network links between a customer geographical location and one or more web site data network geographical locations, and size of a click stream for arriving at one or more web site data network geographical locations". In rejecting Claim 10, the Examiner states that *Wu* teaches a system that determines customer's navigational path through websites or web pages by calculating the amount of links by task, site and speed of search results in order to predict if an increase in a customer's purchase rate was a result of an improvement in the navigational path, citing *Wu's* teaching at column 18, table B and column 36, lines 24-30.

Appellants respectfully urge that such link calculation allegation does not address the specific claimed feature recited in Claim 10 pertaining to *frequency distributions of both number of data links and size of a click stream*. Thus, a proper prima facie showing of obviousness has not been established with respect to the claimed frequency distribution feature recited in Claim 10.

Still further, and even assuming arguendo that the cited reference teaches a frequency distribution of both number of data links and size of a click stream feature (which it does not), this still does not establish any teaching or suggestion of the specific claimed feature that the first and second *statistical distributions that are used to compare to a statistical distribution of a customer* database *to determine if at least one of the training data set and the testing data set are geographically representative of a customer population represented by the customer database* are themselves 'frequency distributions of number of data network links between a customer geographical location and one or more web site data network geographical locations, and size of a click stream for arriving at one or more web site data network geographical locations'. Simply put, even if *Wu* did teach a frequency distribution of both number of data links and size of a click stream feature (which it does not), the existence of such feature does not establish a teaching/suggestion of the *specific use* of such frequency distributions, as expressly recited in Claim 10. Thus, a proper prima facie showing of obviousness has not been established by the Examiner, and accordingly Claim 10 has been erroneously rejected.

C.10. Claims 11, 25 and 38

In addition to the reasons given above with respect to Claim 1 (of which Claim 11 depends upon), Appellants further urge error in the rejection of Claim 11, as such claim recites "wherein comparing at least one of the first statistical distribution and the second statistical distribution to a statistical distribution of a customer database includes: generating a composite data set from the training data set and the testing data set; and generating a composite statistical distribution from the composite data set that was generated from the training data set and the testing data set". As can be seen, a composite data set is generated *from both* the training data set and the testing data set, and a composite statistical distribution is generated from this (generated) composite data set. In rejecting Claim 11, the Examiner cites *Menon's* teaching at column 4, lines 1-15 are teaching the generation of both of these items ((1) a composite data set and (2) a composite statistical distribution). Appellants respectfully urge that this passage describes

combining of two types of observation class histograms together – a voice observation class histogram and a visual data observation class histogram. Importantly, this is *observed data with respect to actual voice and visual data of a user* (col. 13, lines 52-67; Figure 6), and is not the actual testing and training data sets as expressly recited in Claim 11. Thus, a proper prima facie showing of obviousness has not been established by the Examiner, and accordingly Claim 11 has been erroneously rejected.

D. GROUND OF REJECTION 4 (Claims 41-43)

Claims 41-43 stand rejected under 35 U.S.C. § 103(a) as being obvious over *Menon et al.* (U.S. 5,537,488) in view of Appellant's background of the invention and further in view of *Wu* (U.S. 6,741,967).

D.1. Claims 41-43

With respect to Claim 41 (and similarly for Claims 42 and 43), such claim is directed to *predicting customer behavior based on data network geographical influences*, and specifically recites steps of (i) obtaining data network geographical information regarding a plurality of customers, (ii) training a predictive algorithm using the data network geographical information, and (iii) *using the predictive algorithm to predict customer behavior* based on the data network geographical information. Claim 41 is specifically directed to *using actual customer geographic information to train a predictive algorithm*. None of the cited references teach or otherwise suggest *using such customer geographic information to train a predictive algorithm* such that the predictive algorithm can then be used to predict customer behavior *based on the data network geographical information*, for substantially the same reasons as those given above with respect to Claim 1.

In conclusion, Appellants have shown numerous and substantial error in the final rejection of all pending claims in the present application, and respectfully requests that the Board reverse such final rejection of all pending claims.

/Wayne P. Bailey/
Wayne P. Bailey
Reg. No. 34,289
YEE & ASSOCIATES, P.C.
PO Box 802333
Dallas, TX 75380
(972) 385-8777

CLAIMS APPENDIX

The text of the claims involved in the appeal are:

1.   A data processing machine implemented method of selecting data sets for use with a predictive algorithm based on data network geographical information, comprising data processing machine implemented steps of:

generating a first statistical distribution of a training data set;

generating a second statistical distribution of a testing data set;

using the first statistical distribution and the second statistical distribution to identify a discrepancy between the first statistical distribution and the second statistical distribution with respect to the data network geographical information by comparing at least one of the first statistical distribution and the second statistical distribution to a statistical distribution of a customer database to determine if at least one of the training data set and the testing data set are geographically representative of a customer population represented by the customer database;

modifying selection of entries in one or more of the training data set and the testing data set based on the discrepancy between the first statistical distribution and the second statistical distribution; and

using the modified selection of entries by the predictive algorithm.

2.   The method of claim 1, wherein the first statistical distribution and the second statistical distribution are distributions of a number of data network links from a customer data network geographical location to a web site data network geographical location.

3.   The method of claim 1, wherein the first statistical distribution and the second statistical distribution are distributions of a size of a click stream for arriving at a web site data network geographical location.

4.   The method of claim 1, wherein comparing the first statistical distribution and the second statistical distribution includes comparing one or more of a mean, mode, and standard deviation of the first statistical distribution to one or more of a mean, mode, and standard deviation of the second statistical distribution.

5.   The method of claim 1, wherein the first statistical distribution and the second statistical distribution are distributions of a weighted data network geographical distance between a customer data network geographical location and a web site data network geographical locations.

6.   The method of claim 1, wherein the first statistical distribution and the second statistical distribution are distributions of a weighted click stream for arriving at a web site data network geographical locations.

7.   The method of claim 1, wherein modifying selection of entries in one or more of the training data set and the testing data set includes generating recommendations for improving selection of entries in one or more of the training data set and the testing data set, and wherein the method of claim 1 further comprises re-generating at least one of the first statistical distribution and the second statistical distribution based upon the recommendations.

8.   The method of claim 1, wherein the training data set and the testing data set are selected from a customer information database comprising information with respect to customers who have purchased any of goods and services over a data network, wherein the data network geographic information pertains to geographic information of the data network.

10.   The method of claim 1, wherein the first statistical distribution and second statistical distribution are frequency distributions of number of data network links between a customer geographical location and one or more web site data network geographical locations, and size of a click stream for arriving at one or more web site data network geographical locations.

11.   The method of claim 1, wherein comparing at least one of the first statistical distribution and the second statistical distribution to a statistical distribution of a customer database includes:
        generating a composite data set from the training data set and the testing data set; and
        generating a composite statistical distribution from the composite data set that was generated from the training data set and the testing data set.

12.   The method of claim 1, wherein modifying selection of entries in one or more of the training data set and the testing data set includes changing one of a random selection algorithm and a seed value for the random selection algorithm, and then re-comparing the first statistical distribution and the second statistical distribution.

13.   The method of claim 1, wherein using the modified selection of entries by the predictive algorithm includes training the predictive algorithm using at least one of the training data set and the testing data set if the discrepancy is within a predetermined tolerance.

14.   The method of claim 13, wherein the predictive algorithm is a discovery based data mining algorithm.

15.   An apparatus for selecting data sets for use with a predictive algorithm based on data network geographical information, comprising:

a statistical engine;

a comparison engine coupled to the statistical engine, wherein the statistical engine generates a first statistical distribution of a training data set and a second distribution of a testing data set, the comparison engine uses the first statistical distribution and the second distribution to identify a discrepancy between the first statistical distribution and the second distribution with respect to the data network geographical information by comparing at least one of the first statistical distribution and the second statistical distribution to a statistical distribution of a customer database to determine if at least one of the training data set and the testing data set are geographically representative of a customer population represented by the customer database, modifies selection of entries in one or more of the training data set and the testing data set based on the discrepancy between the first statistical distribution and the second distribution, and provides the modified selection of entries for use by the predictive algorithm; and

a predictive algorithm device that uses the modified selection of entries and the predictive algorithm.

16. The apparatus of claim 15, wherein the first statistical distribution and the second statistical distribution are distributions of a number of data network links from a customer data network geographical location to a web site data network geographical location.

17. The apparatus of claim 15, wherein the first statistical distribution and the second statistical distribution are distributions of a size of a click stream to arrive at a web site data network geographical location.

18. The apparatus of claim 15, wherein the comparison engine compares the first statistical distribution and the second statistical distribution by comparing one or more of a mean, mode, and standard deviation of the first statistical distribution to one or more of a mean, mode, and standard deviation of the second statistical distribution.

19. The apparatus of claim 15, wherein the first statistical distribution and the second statistical distribution are distributions of a weighted number of data network links between a customer data network geographical location and a web site data network geographical location.

20. The apparatus of claim 15, wherein the first statistical distribution and the second statistical distribution are distributions of a weighted size of a click stream to arrive at a web site data network geographical location.

21. The apparatus of claim 15, wherein the comparison engine modifies selection of entries in one or more of the training data set and the testing data set by generating recommendations for

improving selection of entries in one or more of the training data set and the testing data set, and wherein the statistical engine re-generates at least one of the first statistical distribution and the second statistical distribution based upon the recommendations.

22.   The apparatus of claim 15, further comprising a training data set/testing data set selection device that selects the training data set and the testing data set from a customer information database comprising information with respect to customers who have purchased any of goods and services over a data network, wherein the data network geographic information pertains to geographic information of the data network.

24.   The apparatus of claim 15, wherein the first statistical distribution and second statistical distribution are frequency distributions of a number of data network links between a customer data network geographical location and one or more web site data network geographical locations, and a size of a click stream to arrive at one or more web site data network geographical locations.

25.   The apparatus of claim 15, wherein the comparison engine compares at least one of the first statistical distribution and the second statistical distribution to a statistical distribution of a customer database by:

    generating a composite data set from the training data set and the testing data set; and

    generating a composite statistical distribution from the composite data set that was generated from the training data set and the testing data set.

26. The apparatus of claim 15, wherein the comparison engine modifies selection of entries in one or more of the training data set and the testing data set by changing one of a random selection algorithm and a seed value for the random selection algorithm, and then re-comparing the first statistical distribution and the second statistical distribution.

27. The apparatus of claim 15, wherein the predictive algorithm device is trained using at least one of the training data set and the testing data set if the discrepancy is within a predetermined tolerance.

28. The apparatus of claim 27, wherein the predictive algorithm is a discovery based data mining algorithm.

29. A computer program product in a computer readable medium comprising instructions for enabling a data processing machine to select data sets for use with a predictive algorithm based on data network geographical information, comprising:

first instructions for generating a first statistical distribution of a training data set;

second instructions for generating a second statistical distribution of a testing data set;

third instructions for using the first statistical distribution and the second statistical distribution to identify a discrepancy between the first statistical distribution and the second statistical distribution with respect to the data network geographical information by comparing at least one of the first statistical distribution and the second statistical distribution to a statistical

distribution of a customer database to determine if at least one of the training data set and the testing data set are geographically representative of a customer population represented by the customer database;

fourth instructions for modifying selection of entries in one or more of the training data set and the testing data set based on the discrepancy between the first statistical distribution and the second statistical distribution; and

fifth instructions for using the modified selection of entries by the predictive algorithm.

30.   The computer program product of claim 29, wherein the first statistical distribution and the second statistical distribution are distributions of a number of data network links from a customer data network geographical location to a web site data network geographical location.

31.   The computer program product of claim 29, wherein the first statistical distribution and the second statistical distribution are distributions of a size of a click stream to arrive at a web site data network geographical location.

32.   The computer program product of claim 29, wherein the third instructions for comparing the first statistical distribution and the second statistical distribution include instructions for comparing one or more of a mean, mode, and standard deviation of the first statistical distribution to one or more of a mean, mode, and standard deviation of the second statistical distribution.

33.   The computer program product of claim 29, wherein the first statistical distribution and the second statistical distribution are distributions of a weighted number of data network links between a customer data network geographical location and a web site data network geographical location.

34.   The computer program product of claim 29, wherein the first statistical distribution and the second statistical distribution are distributions of a weighted size of a click stream to arrive at a web site data network geographical location.

35.   The computer program product of claim 29, wherein the fourth instructions for modifying selection of entries in one or more of the training data set and the testing data set include instructions for generating recommendations for improving selection of entries in one or more of the training data set and the testing data set, and wherein the computer program product claim 29 further comprises instructions for re-generating at least one of the first statistical distribution and the second statistical distribution based upon the recommendations.

37.   The computer program product of claim 29, wherein the first statistical distribution and second statistical distribution are frequency distributions of a number of data network links between a customer data network geographical location and one or more web site data network geographical locations, and a size of a click stream to arrive at one or more web site data network geographical locations.

38. The computer program product of claim 29, wherein the fifth instructions include:

instructions for generating a composite data set from the training data set and the testing data set; and

instructions for generating a composite distribution from the composite data set that was generated from the training data set and the testing data set.

39. The computer program product of claim 29, wherein the fourth instructions for modifying selection of entries in one or more of the training data set and the testing data set include instructions for changing one of a random selection algorithm and a seed value for the random selection algorithm, and then re-comparing the first statistical distribution and the second statistical distribution.

40. The computer program product of claim 29, wherein the fifth instructions include instructions for training the predictive algorithm using at least one of the training data set and the testing data set if the discrepancy is within a predetermined tolerance.

41. A data processing machine implemented method of predicting customer behavior based on data network geographical influences, comprising data processing machine implemented steps of:

obtaining data network geographical information regarding a plurality of customers, the data network geographic information comprising frequency distributions of both (i) number of

data network links between a customer geographical location and one or more web site data

network geographical locations, and (ii) size of a click stream for arriving at the one or more web

site data network geographical locations;

training a predictive algorithm using the data network geographical information; and

using the predictive algorithm to predict customer behavior based on the data network

geographical information.

42.   An apparatus for predicting customer behavior based on data network geographical

influences, comprising:

means for obtaining data network geographical information regarding a plurality of

customers, the data network geographic information comprising frequency distributions of both

(i) number of data network links between a customer geographical location and one or more web

site data network geographical locations, and (ii) size of a click stream for arriving at the one or

more web site data network geographical locations;

means for training a predictive algorithm using the data network geographical

information; and

means for using the predictive algorithm to predict customer behavior based on the data

network geographical information.

43.   A computer program product in a computer readable medium comprising instructions for

enabling a data processing machine to predict customer behavior based on data network

geographical influences, comprising:

first instructions for obtaining data network geographical information regarding a plurality of customers, the data network geographic information comprising frequency distributions of both (i) number of data network links between a customer geographical location and one or more web site data network geographical locations, and (ii) size of a click stream for arriving at the one or more web site data network geographical locations;

second instructions for training a predictive algorithm using the data network geographical information; and

third instructions for using the predictive algorithm to predict customer behavior based on the data network geographical information.

EVIDENCE APPENDIX

There is no evidence to be presented.

RELATED PROCEEDINGS APPENDIX

There are no related proceedings.